# SpaceX First Launch

Will the launch be successful?

# Executive Summary

- In this project, we will predict if the Falcon 9 first stage will land successfully.
- We perform data collection and data wrangling
- We perform exploratory data analysis with SQL and Pandas to understand the data.
- We build interactive maps with Folium to visualize the data.
- We build predictive models using various machine learning algorithms.
- The best accuracy was 84.82%% on the test set.

# Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Therefore if we can determine if the first stage will land, we can determine the cost of a launch.
- This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

# Data Collection & Data Wrangling

- We gathered data from the SpaceX API.

- We web scrapped Falcon 9 launch records with BeautifulSoup.

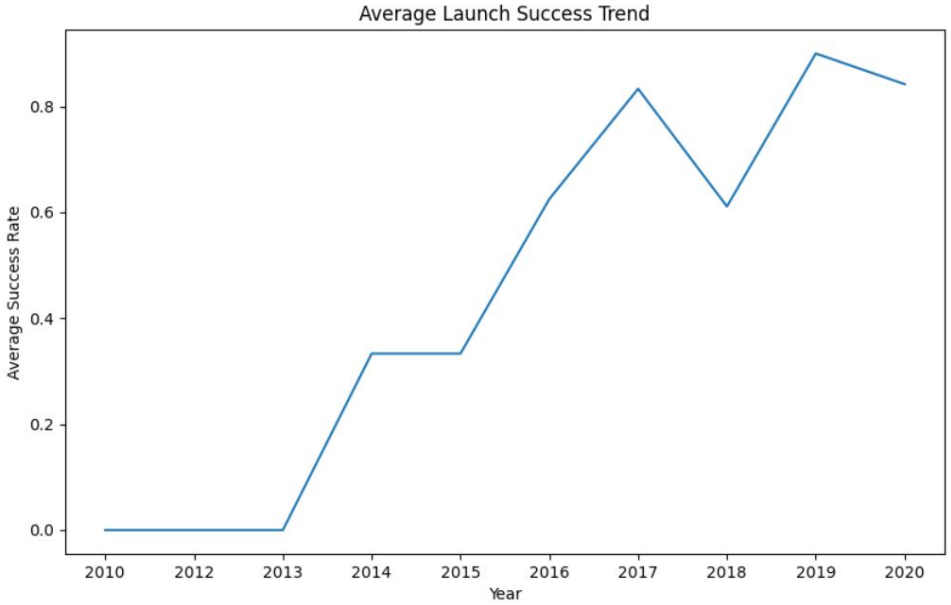
- Extract a Falcon 9 launch records HTML tabl from Wikipedia.

- Parse the table and convert it into a Pandas dataframe.

| | Flight No. | Launch site | Payload | Payload mass | Orbit | Customer | Launch outcome | Version Booster | Booster landing | Date | Time |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | CCAFS | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success\n | F9 v1.0B0003.1 | Failure | 4 June 2010 | 18:45 |
| **1** | 2 | CCAFS | Dragon | 0 | LEO | NASA | Success | F9 v1.0B0004.1 | Failure | 8 December 2010 | 15:43 |
| **2** | 3 | CCAFS | Dragon | 525 kg | LEO | NASA | Success | F9 v1.0B0005.1 | No attempt\n | 22 May 2012 | 07:44 |
| **3** | 4 | CCAFS | SpaceX CRS-1 | 4,700 kg | LEO | NASA | Success\n | F9 v1.0B0006.1 | No attempt | 8 October 2012 | 00:35 |
| **4** | 5 | CCAFS | SpaceX CRS-2 | 4,877 kg | LEO | NASA | Success\n | F9 v1.0B0007.1 | No attempt\n | 1 March 2013 | 15:10 |

Sample Data

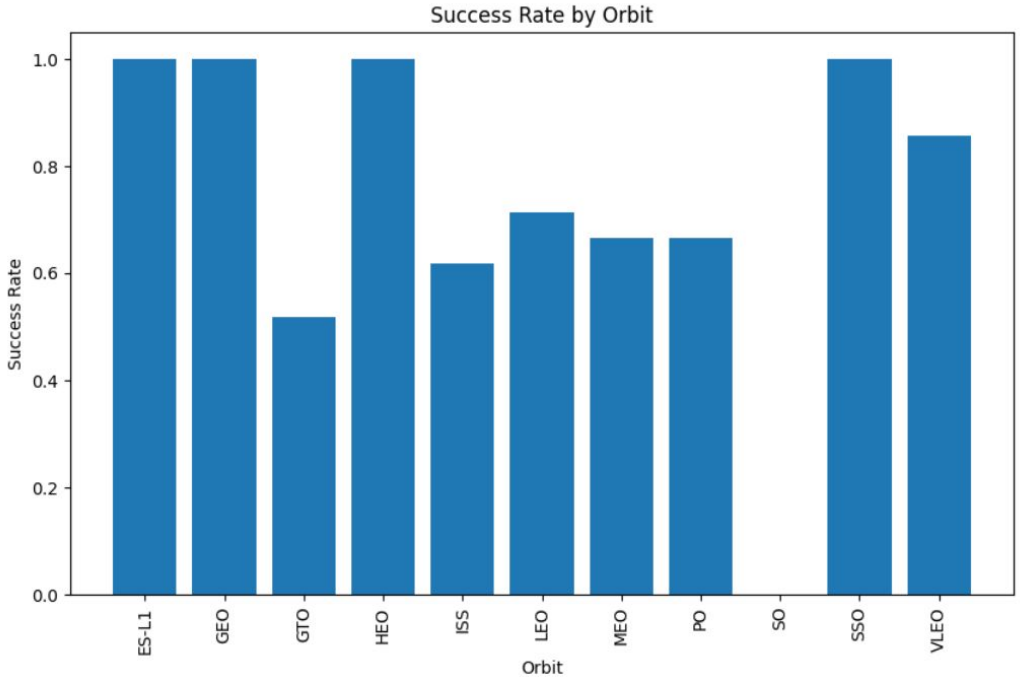
# Exploratory Data Analysis



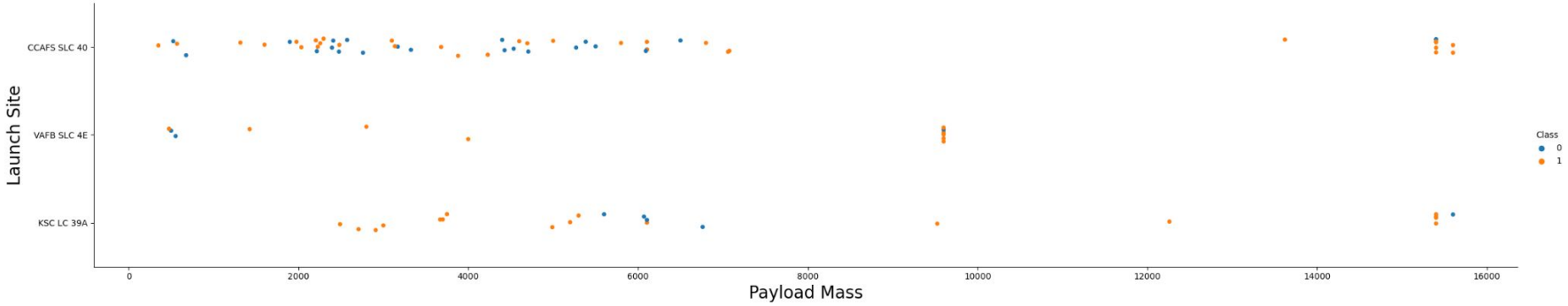

**You can observe that the success rate of launches since 2013 kept increasing till 2020.**

# Exploratory Data Analysis

- **The bar chart shows the success rate of each orbit.**
- **The orbits with the highest success rates are ES-L1, GEO, HEO, and SSO**



# Exploratory Data Analysis




- **Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10000).**

# EDA with SQL

We explore the data using SQL to gain more insight and answer some questions.

```
%%sql
SELECT Customer
FROM SPACEXTBL
WHERE Mission_Outcome = 'Success'
    AND Landing_Outcome LIKE '%drone%'
    AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000);
```

```
 * sqlite:///my_data1.db
Done.
```

| Customer |
|---|
| SES |
| SKY Perfect JSAT Group |
| SKY Perfect JSAT Group |
| SES |
| SES EchoStar |

*List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000*

# EDA with SQL

There are **899** failure missions and **100** successful missions.

```
%%sql
SELECT Mission_Outcome, COUNT(*) as total
FROM SPACEXTBL
GROUP BY Mission_Outcome;
```

```
 * sqlite:///my_data1.db
Done.
```

| Mission_Outcome | total |
|---|---|
| None | 898 |
| Failure (in flight) | 1 |
| Success | 98 |
| Success | 1 |
| Success (payload status unclear) | 1 |

*List the total number of successful and failure mission outcomes*

# EDA with SQL

There are **899** failure missions and **100** successful missions.

```
%%sql
SELECT Mission_Outcome, COUNT(*) as total
FROM SPACEXTBL
GROUP BY Mission_Outcome;
```

```
 * sqlite:///my_data1.db
Done.
```

| Mission_Outcome | total |
|---|---|
| None | 898 |
| Failure (in flight) | 1 |
| Success | 98 |
| Success | 1 |
| Success (payload status unclear) | 1 |

*List the total number of successful and failure mission outcomes*

# EDA with SQL

There are **4** unique launch sites in the dataset.

```
%%sql
SELECT DISTINCT("Launch_Site")
FROM SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

| Launch_Site |
| --- |
| CCAFS LC-40 |
| VAFB SLC-4E |
| KSC LC-39A |
| CCAFS SLC-40 |
| None |

*Display the names of the unique launch sites in the space mission*

# EDA with SQL

Booster versions that have carried the biggest payload.

```
%%sql
SELECT Booster_Version
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
```

```
 * sqlite:///my_data1.db
Done.
```

]:

| Booster_Version |
|---|
| F9 B5 B1048.4 |
| F9 B5 B1049.4 |
| F9 B5 B1051.3 |
| F9 B5 B1056.4 |
| F9 B5 B1048.5 |
| F9 B5 B1051.4 |
| F9 B5 B1049.5 |
| F9 B5 B1060.2 |
| F9 B5 B1058.3 |
| F9 B5 B1051.6 |
| F9 B5 B1060.3 |

*List the names of the booster_versions which have carried the maximum payload mass*

# Predictive Analysis

We build models using the following machine learning algorithms and evaluate the performance of each using the accuracy metric.

- Logistic Regression
- Support Vector Machines
- Decision Tree
- KNN

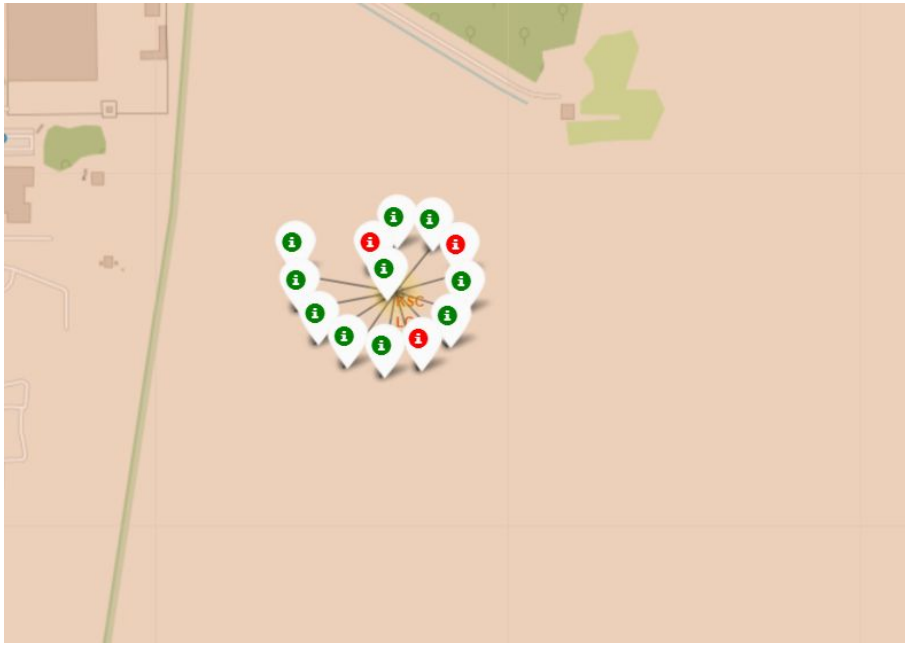
The GridSearchCV algorithm is used to find the optimum hyperparameters for each machine learning algorithm.

# Interactive Folium Map


Map showing the various launch sites

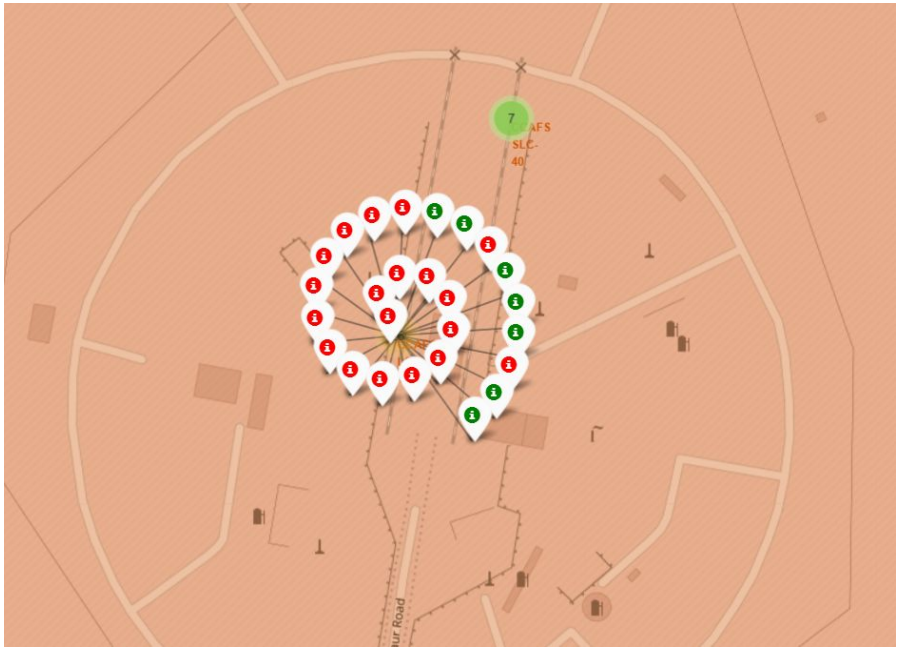
# Interactive Folium Map

The RSC-LC launch site has had **10** successful launches out of 13 launches.



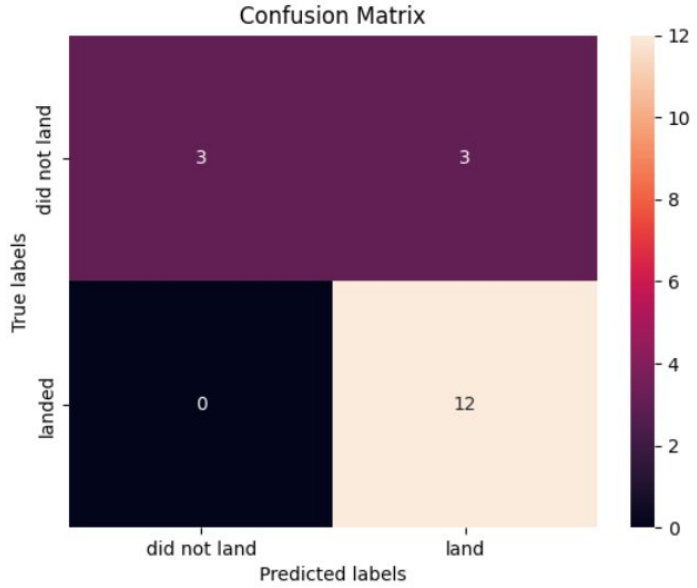
# Interactive Folium Map

The CCAFS launch site has had **7** successful launches out of 26 launches.



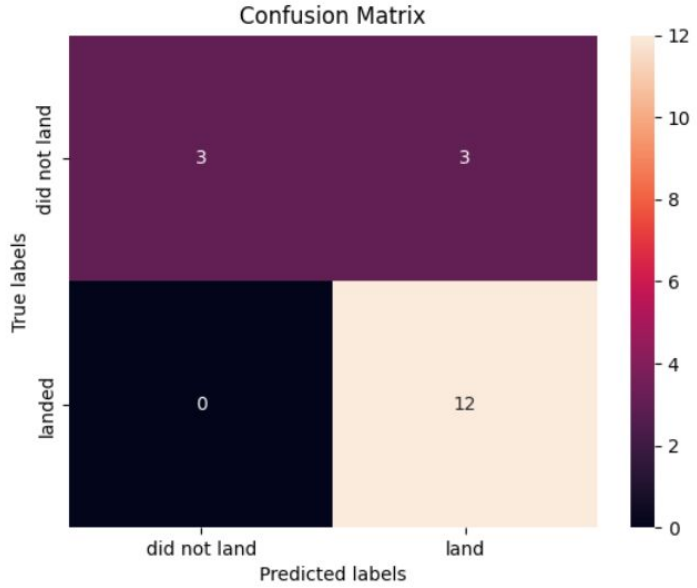
# Predictive Analysis Result

- The **logistic regression** model had a train accuracy of **84.82%** and a test accuracy of 83.33%**.**
- It correctly predicted **12** successful launches and **3** unsuccessful launches.
- It misclassified 3 launches that failed as successful.



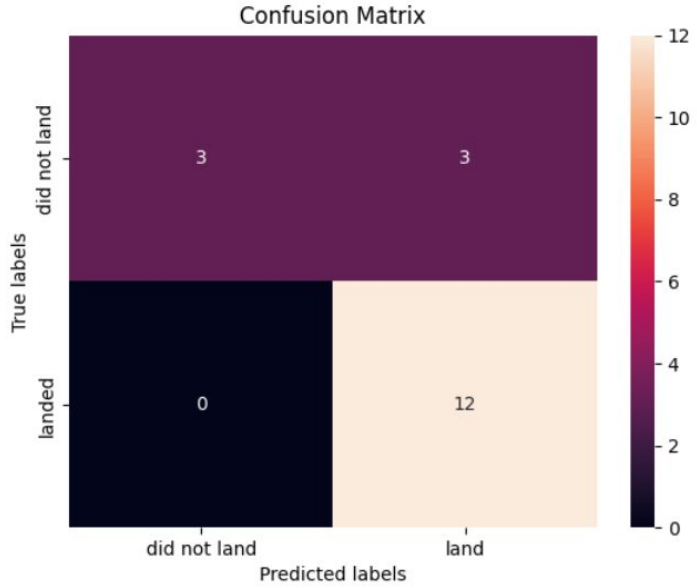
# Predictive Analysis Result

- The **SVM** model had a train accuracy of **84.82%** and a test accuracy of 83.33%**.**
- It correctly predicted **12** successful launches and **3** unsuccessful launches.
- It misclassified **3** launches that failed as successful.



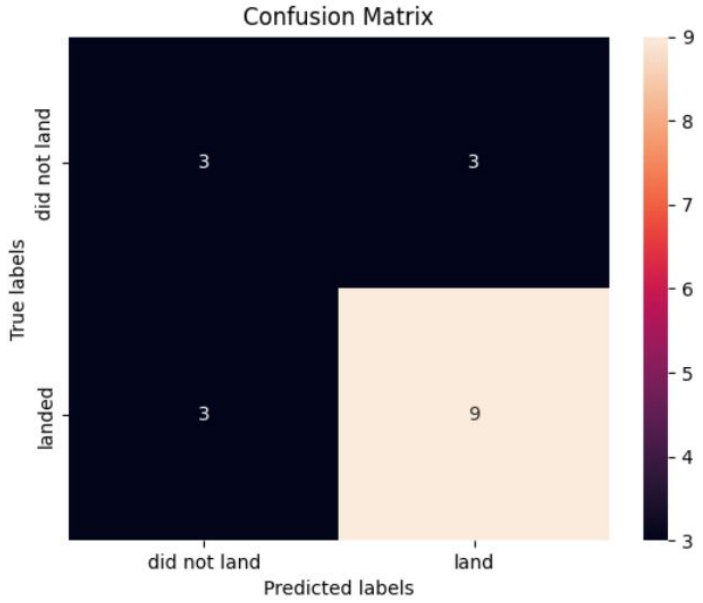
# Predictive Analysis Result

- The **KNN** model had a train accuracy of **84.82%** and a test accuracy of 83.33%**.**
- It correctly predicted **12** successful launches and **3** unsuccessful launches.
- It misclassified **3** launches that failed as successful.



# Predictive Analysis Result

- The **Decision Tree** model had a train accuracy of **88.93%** and a test accuracy of **66.67%.**
- It correctly predicted **9** successful launches and **3** unsuccessful launches.
- It misclassified **3** launches that failed as successful.
- It misclassified **3** launches that succeeded as failures.



# Conclusion

- Most of the launches from the CCAFS site were unsuccessful. Further investigation is needed.
- Launches aimed toward the ES-L1, GEO, HEO, and SSO orbits have a higher chance of being successful.
- The machine learning model was successful at predicting rockets that landed successfully. However, more data is needed to improve the performance of the machine learning model.